

De la segmentation dans les tweets : signes de ponctuation, connecteurs, émoticônes et émojis

Segmentation devices in tweets: punctuation marks, connectives, emoticons and emojis

Jean-Philippe Magué, Nathalie Rossi-Gensane and Pierre Halté

**Electronic version**

URL: <http://journals.openedition.org/corpus/4619>

DOI: 10.4000/corpus.4619

ISSN: 1765-3126

Publisher

Bases ; corpus et langage - UMR 6039

Electronic reference

Jean-Philippe Magué, Nathalie Rossi-Gensane and Pierre Halté, « De la segmentation dans les tweets : signes de ponctuation, connecteurs, émoticônes et émojis », *Corpus* [Online], 20 | 2020, Online since 24 January 2020, connection on 08 September 2020. URL : <http://journals.openedition.org/corpus/4619> ; DOI : <https://doi.org/10.4000/corpus.4619>

De la segmentation dans les tweets : signes de ponctuation, connecteurs, émoticônes et émojis¹

Segmentation devices in tweets: punctuation marks, connectives, emoticons and emojis

Jean-Philippe MAGUÉ*, Nathalie ROSSI-GENSANE**
et Pierre HALTÉ***

* ICAR (UMR 5191) et ENS de Lyon ; ** ICAR (UMR 5191)
et Université Lumière Lyon 2 ; *** EDA (EA 4071) et
Université de Paris

Résumé

Dans cet article, nous appuyant sur un corpus de 3 444 075 tweets correspondant à 44 107 210 tokens (mots, signes de ponctuation, émojis, émoticônes, etc.) recueillis en décembre 2016, nous nous intéressons aux procédés de segmentation à l'œuvre dans les tweets. Après avoir évoqué certaines caractéristiques de ces écrits particuliers, nous rappelons les procédés généraux de segmentation à l'écrit : les signes de ponctuation et les connecteurs. Nous nous penchons ensuite sur la segmentation opérée dans les tweets par ces deux procédés généraux. Enfin, nous montrons que les émoticônes et les émojis constituent des procédés spécifiques permettant de diversifier les stratégies de segmentation des utilisateurs de tweets (et d'autres écrits numériques, tels les SMS et les courriels).

Mots-clés

segmentation, tweets, signes de ponctuation, connecteurs, émoticônes, émojis

¹ Étude réalisée dans le cadre du projet SoSweet soutenu par l'ANR (ANR-15-CE38-0011-03).

Abstract

In this paper, relying on a corpus of 3,444,075 tweets corresponding to 44 107 210 tokens (words, signs of punctuation, emojis, emoticons, etc.) collected in December 2016, we focus on segmentation processes at work in tweets. After mentioning some characteristics of these particular writings, we review the general segmentation processes in writing, punctuation and connectors. We then look at how these processes operate in tweets. Finally, we show that emoticons and emojis are specific processes allowing users to diversify their segmentation strategies (and other digital writings, such as SMS and email).

Keywords

segmentation processes, tweets, punctuation marks, connectives, emoticons, emojis

Introduction

Les tweets font partie de ces genres liés à l'apparition de nouvelles technologies (par exemple, les SMS, les tchats, les courriels), qui, « manifestation[s] de l'immédiat dans le graphique », « viennent introduire [du] bougé dans les paramètres fondamentaux de la communication par oral et par écrit » alors que, « jusqu'à il y a un siècle et demi environ, l'oral et l'écrit se présentaient comme des entités stabilisées dans quelques propriétés fondamentales » (Gadet, 2007-2008 : 135 et 136).

Dans cet article, nous appuyant sur un corpus de 3 444 075 tweets correspondant à 44 107 210 tokens (mots, signes de ponctuation, émojis, émoticônes, etc.) recueillis en décembre 2016, nous nous intéressons aux procédés de segmentation à l'œuvre dans les tweets. Après avoir évoqué certaines caractéristiques de ces écrits particuliers, nous rappelons les procédés généraux de segmentation à l'écrit : les signes de ponctuation et les connecteurs. Nous nous penchons ensuite sur la segmentation opérée dans les tweets par ces deux procédés généraux. Enfin, nous nous interrogeons sur le rôle des émoticônes et des émojis en tant que procédés de segmentation

spécifiques aux tweets (à côté d'autres écrits numériques, tels les SMS et les courriels).

1. Les tweets : considérations générales

À la suite de Söll (1974), Koch & Oesterreicher (2001) différencient nettement, dans tout énoncé, l'aspect médial, lié au canal, généralement oral ou écrit, et l'aspect conceptionnel, pour sa part lié à la proximité ou la distance communicatives. Dans les tweets, écrits sur le plan médial, ressort particulièrement « le caractère scalaire » (Koch & Oesterreicher 2001 : 586) du plan conceptionnel. Certains « paramètres de la communication » (Koch & Oesterreicher 2001 : 586) sont en effet clairement déterminés, tels le « détachement actionnel et situationnel », le « détachement référentiel de la situation », la « séparation spatio-temporelle » (tout au moins la séparation spatiale, une coprésence temporelle étant possible) et, sans doute, une « coopération communicative minimale » et une « liberté thématique ». En revanche, d'autres paramètres sont soumis à une forte gradation interne : les tweets livrent des messages susceptibles de relever de la « communication privée » ou, s'il s'agit de tweets collectifs, « publique », avec « interlocuteur intime » ou « inconnu », empreints d'une « émotionnalité forte » ou « faible », s'inscrivant dans un « monologue » ou dans un « dialogue » (mais, dans ce dernier cas, généralement, en différé), la communication pouvant être « spontanée » comme « préparée ». Les tweets sont donc dotés d'un « relief conceptionnel » (Koch & Oesterreicher 2001 : 586) à géométrie variable. De ce fait, si l'on convoque la notion de genre textuel, définie comme une « corrélation entre fonctionnements linguistiques et situation extralinguistique » (Condamines 2006 : 633), sans doute convient-il de parler de genres, au pluriel, des tweets.

Les tweets peuvent de plus être considérés comme relevant de ce que l'on appelle parfois les genres brefs (voir notamment Roukhomovsky (2001)). Jusqu'en novembre 2017 (date postérieure au recueil de notre corpus qui a eu lieu en décembre 2016), leur longueur était limitée à 140 caractères (elle peut aujourd'hui aller jusqu'à 280). Les formes brèves que

constituent les tweets sont à elles seules un texte séparé (ce qui ne serait pas nécessairement le cas pour des genres brefs tels que les maximes, susceptibles de se trouver au sein d'un autre texte, par exemple en tant que morale d'une histoire), matériellement contraint par la limite du nombre de caractères, d'une façon, par certains côtés, comparable à l'épigramme qui « ne devait contenir plus de vers qu'il s'en pouvait écrire dessus un portail » (Sébillot 1548).

Enfin, dans cette réflexion sur la segmentation dans les tweets, il est important de souligner qu'en pays de littérature, que Gadet (2007 : 45) décrit comme « la pratique généralisée de la lecture-écriture, les effets d'une culture de l'écrit sur les énoncés, les pratiques, attitudes et représentations, pour un locuteur ou une communauté de locuteurs », la standardisation s'applique d'abord aux genres de l'écrit et, surtout, à ceux relevant de l'écrit littéraire, dans la mesure où ce sont « les meilleurs écrivains qui [sont considérés comme] fourniss[a]nt les modèles les plus sûrs de la Norme » (Lodge 2011 : 87). On peut ainsi s'attendre à ce que les tweets tendent à être moins normés que d'autres écrits, en particulier sur le plan de la segmentation qui nous intéresse ici.

2. Les tweets du corpus

Notre corpus a été acquis auprès de l'entreprise Gnip, et correspond à 10% des tweets produits en décembre 2016 dans les fuseaux horaires GMT et GMT+1 et identifiés par Twitter comme étant en français. Afin d'éliminer les tweets créés automatiquement par des robots, nous avons filtré notre corpus en fonction du client (le logiciel) utilisé pour produire les tweets. Nous avons écarté les clients contenant le mot *bot* (pour robot) ou comptant moins de 1000 tweets. Les tweets automatiques étant très standardisés, nous avons calculé l'écart-type de la longueur des tweets pour chaque client. Ceux pour lesquels cet écart-type est inférieur à 6, c'est-à-dire présentant une faible diversité, ont été également éliminés. Notre corpus final compte 34 clients différents. Les cinq plus fréquents (Twitter for iPhone, Twitter for Android, Twitter Web Client, TweetDeck et Twitter for iPad), qui représentent 97.7% des tweets de notre corpus, sont clairement des logiciels conçus pour être utilisés par des humains

et non des robots. Le corpus initial se montait à 4 243 789 tweets ; le corpus filtré en comprend 3 444 075. Ce corpus constitue donc un échantillon substantiel des énoncés produits par les utilisateurs francophones de Twitter en Europe (et, dans une moindre mesure, en Afrique). Sans nier la diversité des productions qu'il contient (analysée par ailleurs, voir par exemple Abitbol *et al.* (2015)), notre étude se concentre sur les procédés de segmentation à l'échelle de l'ensemble de ce corpus (Bolander & Locher 2014).

Dans le corpus, nous avons identifié et décompté les séquences (éventuellement constituées d'un seul élément) de signes de ponctuation, d'émoticônes ou d'emojis avec des expressions régulières. Dans l'exemple ci-dessous, :(((compte pour une séquence d'émoticônes, ?? pour une séquence de signes de ponctuation, et 😊😊😊😊 pour une séquence d'emojis :

(1) @X je sais jamais les anniversaires en vrai :(((
mais c'est le geste qui compte n'est ce pas??
MDRRRR😊😊😊😊

Dans le cadre de cet article, nous nous sommes limités aux signes de ponctuation suivants : la virgule, le point, le point d'interrogation, le point d'exclamation et les points de suspension. Les emojis et les émoticônes retenus sont, quant à eux, donnés en annexe. Par ailleurs, étant donnée la diversité des modes de présentation des tweets selon le logiciel utilisé, nous avons délibérément opté pour une mise en forme de nos exemples qui ne préserve que leur contenu linguistique, sans essayer de mimer l'interface d'un logiciel arbitraire. La graphie d'origine a été conservée. De plus, comme il apparaît dans l'exemple ci-dessus, nous avons anonymisé les mentions aux utilisateurs en les remplaçant systématiquement par @X.

3. Les procédés généraux de segmentation à l'écrit : les signes de ponctuation et les connecteurs

Tout texte, pour être appréhendé, doit être préalablement découpé. Ce rôle, tout au moins dans les écrits normés, est

généralement dévolu à la ponctuation, définie de manière très large comme l'« ensemble des signes graphiques non alphabétiques utilisés dans un texte pour noter les rapports syntaxiques entre les divers éléments de la phrase ou de la proposition, les rapports avec le sens, les idées du texte, les variations d'ordre affectif (intonation, rythme, mélodie de la phrase) » (TLFi, CNRTL). En particulier, Dürrenmatt (2015 : 23) envisage, parmi les « opérations proposées au lecteur par la ponctuation », celle de l'« agence[ment] », grâce à la virgule, au point-virgule, au deux-points, au point et, dans une moindre mesure, au point d'interrogation, au point d'exclamation, aux points de suspension : « le ponctuant rend visibles des unités et permet de les interpréter pour elles-mêmes et dans leurs rapports les unes avec les autres ». Néanmoins, cette opération d'agencement n'est pas réservée aux seuls signes de ponctuation, comme le souligne également Dürrenmatt (2015 : 23), car, dans la mesure où les unités susmentionnées « sont de même niveau : la ponctuation participe [...] de la jonction ». Dürrenmatt (2015 : 23) signale ainsi que « les ponctuants peuvent [...] accompagner des connecteurs ». Outre le possible cumul entre signes de ponctuation et connecteurs, il est important de préciser que les connecteurs sont dotés de la capacité d'assurer seuls cette mission d'agencement et que, par conséquent, signes de ponctuation et connecteurs doivent être vus comme agissant à l'écrit de manière complémentaire. De plus, dans les tweets, comme dans des écrits d'autres genres liés aux technologies numériques, tels les SMS ou les courriels, le scripteur a la possibilité de recourir à des émoticônes et à des émojis, susceptibles d'apparaître aux mêmes endroits de la chaîne écrite (voir *infra*, partie 5). De par la complémentarité entre signes de ponctuation et connecteurs, que l'on peut aussi envisager d'étendre ici aux émoticônes et aux émojis, il semble pertinent d'utiliser la notion de segmentation, plus large que celle de ponctuation.

La segmentation, « division en segments » d'après le TLFi (CNRTL), sera entendue comme la démarcation visible, matérialisée à l'aide de bornes de diverses sortes (signes de ponctuation, connecteurs, etc., mais aussi, comme il sera discuté

infra, émoticônes et émojis), d'un texte en unités. Toutefois, selon l'unité adoptée, cette démarcation peut se situer à plusieurs niveaux. Dans cet article, nous privilégierons l'unité maximale syntaxique, que nous dénommons phrase « syntaxique » (voir par exemple Paolacci & Rossi-Gensane (2012), à partir de la notion de phrase chez Feuillard (1989)) et qui est loin de toujours correspondre à la phrase telle qu'elle est conçue traditionnellement, c'est-à-dire comme étant délimitée à l'initiale par une majuscule et en finale par un point (ou tout autre signe de ponctuation dite forte, comme le point d'exclamation, le point d'interrogation, les points de suspension). La phrase « syntaxique » est définie comme constituée de l'ensemble des éléments reliés par des rapports de dépendance à un même élément central, qui n'est pas nécessairement verbal (comme, par exemple, dans *Quelle belle journée !*, où l'élément central est *journée*). Lorsque l'élément central est verbal, la phrase « syntaxique » équivaut généralement à une proposition indépendante ou à une proposition principale accompagnée de sa (ou ses) proposition(s) subordonnée(s). Ainsi, dans le cas des propositions traditionnellement dites coordonnées, *Marie déteste la soupe et elle adore le chocolat.*, là où l'on voit habituellement une seule phrase déterminée à partir d'un critère typographique (majuscule-point), nous dégagerons, à partir d'un critère relationnel, plusieurs phrases « syntaxiques », en l'occurrence deux (d'une part, *Marie déteste la soupe*, d'autre part, *et elle adore le chocolat*, les éléments du premier ensemble n'entretenant aucune relation de dépendance syntaxique avec ceux du second ensemble, et inversement). Dans cet exemple normé, on note qu'il ne manque aucune borne : la première, entre les deux phrases « syntaxiques » (plus précisément, au début de la seconde phrase « syntaxique »), est matérialisée par un connecteur (*et*) ; la seconde, à la fin de la seconde phrase « syntaxique », est matérialisée par un point. Alors que la phrase traditionnelle, « typographique » dans les faits, est certes immédiatement identifiable en réception mais instable en production (si l'on pense aux innombrables façons dont un texte est susceptible d'être ponctué), la phrase « syntaxique », qui s'abstrait de la ponctuation et s'appuie sur un critère syntaxique,

garantit généralement sur ce dernier plan un même « calibrage » tout au long du texte.

La segmentation peut également être prise en compte en deçà de la phrase « syntaxique ». C'est ainsi que, par exemple, à côté d'une virgule séparant deux phrases « syntaxiques », appelée virgule « interphrastique » (et alors dotée d'un rôle de ponctuation forte), nous distinguerons une virgule « intraphrastique » séparant des mots ou des syntagmes². De même, un mot en apparence identique, tel *et*, peut jouer le rôle de connecteur quand il sépare deux phrases « syntaxiques » ou, à la manière de la virgule « intraphrastique », le rôle de coordonnant quand il sépare (tout en les reliant) deux éléments de même fonction.

Enfin, on signalera une différence d'orientation entre signes de ponctuation et connecteurs, particulièrement nette lorsqu'ils ont un rôle de démarcation interphrastique. Alors que les signes de ponctuation clôturent ce qu'ils segmentent, les connecteurs, à l'inverse, l'ouvrent, ce qui rend notamment compte du fait que, sauf exception, un connecteur n'apparaîtra pas en finale de texte et que, généralement, dans les cas où signe(s) de ponctuation et connecteur(s) sont contigus, ceux-là précèdent ceux-ci.

4. La segmentation dans les tweets par la ponctuation et les connecteurs

4.1. La segmentation dans les tweets par la ponctuation

Nous avons identifié 472 séquences de signes de ponctuation différentes, avec 0.84 séquence par tweet en moyenne. Les cinq séquences les plus fréquentes sont constituées d'un seul signe :

, 0.226 par tweet
. 0.181 par tweet
! 0.091 par tweet
... 0.090 par tweet

2 Pour la tradition, qui n'envisage qu'une phrase « graphique », la virgule est par définition nécessairement intraphrastique.

? 0.083 par tweet

Pour ce qui concerne les signes de ponctuation, au premier abord, on note que sont employées dans les tweets davantage de virgules que de points, ce qui peut notamment s'expliquer par le fait que, du point de vue de la segmentation, la virgule est un signe bivalent, potentiellement interphrastique ou intraphrastique, contrairement au point, presque toujours interphrastique. Sans doute en relation à cette bivalence, la virgule est considérée comme un signe particulièrement complexe, voire « le plus difficile de tous les signes de ponctuation » (Popin 1998 : 38 ; voir aussi Catach (1994 : 64-71) ou Drillon (1991 : 145)).

Cette supériorité numérique de la virgule sur le point apparaît tout particulièrement dans des tweets où la virgule est utilisée dans un rôle interphrastique (où elle se substitue, donc, au point) et où, possiblement, le point final est absent, phénomène relativement répandu qui semble indiquer qu'à cet emplacement, le point est jugé redondant. En effet, seulement 6.33% des tweets ont un point final et, plus généralement, seulement 18,6% se terminent par un signe de ponctuation, soit, par ordre décroissant, le point, le point d'interrogation, le point d'exclamation.

- (2) Emma elle connaît trop de mots/expressions françaises, jla soupçonne d'avoir passé les 3/4 de sa vie ici

L'exemple suivant, outre des virgules interphrastiques, comporte également une virgule intraphrastique située entre *mois* et *voir* [*sic*] :

- (3) @X @X Non, je pense qu'il est même sorti y'a moins d'un mois, voir une semaine, il est même pas référencé sur Amazon US

Si l'exemple (4) nécessite une borne interphrastique matérialisée après *sérieux*, il contient une virgule intraphrastique après *hier* :

- (4) Sérieux depuis hier, j'ai des messages d'alerte qu'il n'est pas en état de fonctionner

Les exemples de cette sorte ne sont pas sans évoquer l'usage de la virgule « par défaut » que Bessonnat (1991 : 24) souligne dans une perspective didactique pour les élèves de collège : tout se passe « comme si l'élève, emporté par le flux de l'écriture, craignait de prendre la décision coûteuse du point qui casserait l'inspiration ». Il convient tout de même de signaler, à l'inverse, des cas totalement normés sur le plan de la segmentation, où les tweets sont composés de phrases, à la fois « syntaxiques » et « graphiques », démarquées par une majuscule à l'initiale et un point en finale :

- (5) J'ai pas envie de retourner au collège demain.

- (6) Les pions sont aller chercher les gens de ma classe au foyer. Ils leur ont donné des exercices. J'étais pas là.

Dans d'autres cas, moins normés, le point final est présent, toutefois sans qu'il y ait de majuscule à l'initiale, ou alors c'est la majuscule qui est présente sans le point :

- (7) @X j'ai inventé un mot je crois, c'était de rigueur.

- (8) Mon cousin va rester tout seul en Allemagne

Il est intéressant de remarquer que les écarts par rapport à la norme impliquant les signes de ponctuation et les majuscules relèvent pour les uns et les autres des erreurs à « dominante idéogrammique » (Catach 2003 : 282).

Peut aussi être constatée la supériorité numérique du point d'exclamation sur le point d'interrogation (qui se reproduit pour les séquences à plusieurs éléments, le point d'exclamation dupliqué (0.01050 par tweet) étant par exemple plus fréquent que le point d'interrogation dupliqué (0.00430 par tweet)). Cette

constatation est conforme à la remarque suivante de Dürrenmatt (2015 : 15) :

« S'il [le point d'interrogation] reste obligatoire dans le "bon usage", c'est par volonté d'uniformisation. L'ajout par le scripteur du signe à sa phrase interrogative relève donc du respect de la norme et non d'une nécessité communicationnelle. Il apparaît dès lors comme facilement volatile alors que le point d'exclamation, indice d'une modalité affective moins explicitement visible à travers des modifications morphosyntaxiques, résiste beaucoup mieux. »

La présence d'un point d'interrogation dans l'exemple (9) n'est en effet pas liée à une « nécessité communicationnelle », la modalité interrogative étant déjà indiquée par un élément interrogatif, ici *in situ* (c'est-à-dire en même place que l'élément correspondant dans une phrase assertive) :

(9) @X donc pour vous ça prouve quoi?³ Certainement pas l'absence de conflits d'intérêts! #Filimbi #lucha

En revanche, le point d'interrogation a une valeur distinctive (et significative) pour les interrogatives totales dans les exemples suivants, qui, conformément à une tendance importante de l'oral, ne bénéficient d'aucun marquage morphologique :

(10) @X j'essaye de m'améliorer... T'es forte en classe toi?

(11) @X comme d'habitude , tu passes à la cafetria ?

Dans l'exemple suivant contenant un point d'exclamation, la « modalité affective » qu'il indique est néanmoins « visible à travers des modifications

³ La présence ou l'absence d'un espace avant un signe de ponctuation n'ont pas été commentées.

morphosyntaxiques » (Dürrenmatt 2015 : 15), c'est-à-dire l'adverbe exclamatif *tellement* :

(12) @X un vrai délice, et tellement d'actualité!

Dans de nombreux cas, les points d'exclamation et d'interrogation sont utilisés en séquences de plusieurs éléments, généralement homogènes (s'il s'agit de la répétition du même signe), néanmoins quelquefois hétérogènes⁴ :

(13) @X non je trouve que ton Wolf est bien fait !!!!!
J'aime surtt les yeux !!!!

(14) Je sors du kine, y a une voiture allumé je tourne
la tête et je vois une chèvre dans la voiture... C'est
quoi ce Putain de délire ??

(15) Qui c'est qui taff dans 4h ?!??? C bibi

Si, pour Popin (1998 : 34 et 36), qui s'est surtout penché sur les écrits littéraires, il est « possible de faire des séries de points d'interrogation » et « il n'est pas impossible de répéter les points d'exclamation, voire de combiner un point d'exclamation et un point d'interrogation », pour Dürrenmatt (2015 : 26), en revanche, « il est courant de multiplier les points d'exclamation ou d'interrogation pour signaler l'intensité de la modalisation ». Dürrenmatt (2015 : 100) envisage les emplois de cette sorte comme relevant de la visibilité définie par Anis (1983 : 89), selon laquelle « les formes graphiques [sont] un corps signifiant intégré aux isotopies textuelles ». Dürrenmatt (2015 : 101) évoque également cet usage de la ponctuation déploré par le poète italien Leopardi, où les « idées [sont] représent[ées] par des agencements de signes de ponctuation au lieu [d'être] exprim[ées] de façon complexe par des mots ».

⁴ Par exemple, les séquences homogènes !!!, !!!! et !!!!! sont respectivement utilisées à hauteur de 0.00794, 0.00255 et 0.00095 par tweet ; les séquences homogènes ??? et ???? sont respectivement utilisées à hauteur de 0.00317 et 0.00126 par tweet ; les séquences hétérogènes ?! et !? sont respectivement utilisées à hauteur de 0.00191 et 0.00036 par tweet.

Enfin, pour ce qui concerne les points de suspension, Dürrenmatt (2015 : 94) en souligne le rôle « affectif », « dans la mesure où ils marquent la perte de contrôle, le refus de dire, la difficulté à maîtriser son discours et, par là même, rejoignent l'exclamation » :

(16) Ça me manque le temps ou je préparais mes poissons, je les mettais dans mon sac avec du scotch et je les collais sur les gens à la récré...

(17) @X Pendant ce temps à Taïwan, on peut faire balayer la cours aux fouteurs de merde...

Apparaît aussi un emploi minoritaire du « point multiple » (Drillon 1991 : 136) parfois constitué de deux points (0.01815 par tweet, soit un emploi cinq fois moindre que celui normé des trois points) :

(18) @X même mes parents préfère ma soeur..

Drillon (1991 : 137) évoque l'usage que Françoise Sagan avait souhaité faire de ce « point multiple » à deux points, dans le titre de son roman *Aimez-vous Brahms..*

4.2. La segmentation dans les tweets par les connecteurs

Pour ce qui concerne les connecteurs, nous nous pencherons, dans le cadre de cet article, surtout sur l'élément *mais*, qui, avec *et*, fait partie des connecteurs et coordonnants les plus fréquents (respectivement, 207 270 occurrences et 339 515 occurrences). À la différence de *et* dont les emplois de coordonnant sont moins rares, *mais* est presque toujours connecteur dans les tweets, où il se comporte donc comme une sorte d'hyperconnecteur.

Dans l'exemple (19), conformément à un usage écrit répandu, *mais* exprime une contradiction logique (accompagné, dans la phrase « syntaxique » suivante, par un autre connecteur, *donc*, exprimant, quant à lui, une déduction logique) :

- (19) Je suis motivée mais je vais être traumatisée
donc vous avez intérêt à me soutenir

Souvent, on note un cumul signe de ponctuation (surtout la virgule)/connecteur dans cet ordre, conformément à une tendance extrêmement répandue (dans les cas de contiguïté entre signes de ponctuation et connecteurs, la ponctuation précède le connecteur dans 74478 cas et le suit dans 6988 cas) :

- (20) @X Bonjour, j'ai eu écho comme quoi, on
pouvait acheter des boosters avec des ogrines, mais
je ne trouve pas l'option dans la boutique
- (21) @X S'insurger pour quoi? C'est internet. Je
trouve le post que t'as mis révoltant, mais je vais
pas partir en croisade pour autant.

Toutefois, dans de nombreux cas, *mais* semble indiquer, plutôt que l'« inversion argumentative [formulée dans les travaux d'Oswald Ducrot] », un « [changement] de point de vue sur [l']objet [de discours] » selon les termes de Morel & Danon-Boileau (1998 : 118) caractérisant cet élément à l'oral :

- (22) Visite sur TF1.fr (pas ma faute j'ai cliqué). Vous
entendez pas mais derrière y a en + une vidéo de
pub en autoplay. <http://t.co/ZetwU4VRGv>

Cette valeur est particulièrement nette en début de question, *mais* fonctionnant plus comme une « particule énonciative » (Bouchard 2001) que comme un connecteur :

- (23) Mais c'est quoi la différence entre un pansexuel
et un bisexuel?
- (24) @X mais t'es sérieux à faire des histoires comme
ça !? Ouah mais le gars

Dans l'exemple précédent, on notera en outre une seconde occurrence de *mais* dans un usage exclamatif (malgré l'absence d'un point d'exclamation), qui, d'après Bouchard (2001 : 66), « d'une part manifeste une émotion ressentie individuellement et d'autre part la verbalise, la donne à entendre et à partager à autrui ».

Bouchard (2001 : 71) note des usages variés pour des éléments tels que *mais*, *alors* ou *donc*, entre « purs régulateurs de l'action » dans les oraux polygérés et « instructions d'inférence » dans les écrits monogérés. Cette diversité semble se déployer tout particulièrement dans les tweets, écrits souvent spontanés, de par leur nature hybride.

Outre les connecteurs, d'autres éléments apparaissent comme jouant sous forme de mot(s) un rôle démarcatif dans les tweets. Ainsi, Morel & Danon-Boileau (1998 : 94) évoquent, pour l'oral, des éléments à l'initiale, tels *bon* et, tout au moins dans certains cas, *oui* ou *ouais*, appelés « ligateurs énonciatifs », qui, avec d'autres éléments en finale, appelés « ponctuels », tels *hein* et *quoi*, assurent « la régulation de la coénonciation, qui permet d'explicitier la position de l'énonciateur ». Dans l'exemple (25), de même qu'à l'oral, des ligateurs énonciatifs constituent une borne à gauche (en cela, à la manière des connecteurs) :

- (25) Ouai bon je viens de voir un tweet disant que
l'histoire de la S2 de tokyo ghoul était bien on va
aller se recoucher en faite

Dans l'exemple suivant, on note, à la frontière entre les deux premières phrases « syntaxiques », un cumul virgule/connecteur/ligateur énonciatif/virgule :

- (26) @X Et c'était le cas, mais bon, je ne pense pas
que Oda puisse nous faire une chose pareille,
pendant ce temps, apprécions

Dans les exemples suivants, les ponctuels constituent une borne à droite (en cela, à la manière des signes de

ponctuation, en particulier de ponctuation forte). Selon Morel & Danon-Boileau (1998 : 102), *hein* « participe à la construction d'une convergence de points de vue » :

(27) @X vaux mieux tard que jamais hein

En revanche, *quoi*, toujours selon Morel & Danon-Boileau (1998 : 102), « signifie à autrui qu'on énonce sa position à soi et qu'elle n'est pas soumise à discussion » :

(28) @X ta du manger trop vite lol
ne rien faire le soir du 31 c chaud quoi

On remarquera, en finale de la première phrase « syntaxique », outre la présence d'une interjection acronymique (voir *infra*, partie 5), celle d'un saut de ligne qui fait parfois office de démarcation dans les tweets (comme dans d'autres écrits non normés, tels les textes d'élèves).

4.3. De la sous-segmentation (et de la sur-segmentation) dans les tweets

Rappelons que, dans les écrits non normés, tels les textes d'élèves, la notion de ponctuation appelle en creux celle de sous-ponctuation (Béguelin 2000). Cette sous-ponctuation opère à un plus ou moins grand degré selon que les signes de ponctuation sont totalement absents ou bien seulement manquants à certains endroits où une borne devrait être matérialisée (notamment en fin de phrase « syntaxique »). De même, la notion de segmentation (qui permet de prendre en compte des bornes de diverses natures) peut être envisagée comme appelant en creux celle de sous-segmentation, dans des tweets « écrits au kilomètre » où des signes de ponctuation et/ou des connecteurs sont absents à certains emplacements où une borne matérialisée est obligatoire.

L'exemple (29) se caractérise par une sous-segmentation à un très grand degré de par l'absence totale de ponctuation et de connecteurs (*et* jouant un rôle de coordonnant) :

(29) excellent le projet final de maena et tout j'adore

On note dans l'exemple suivant, par ailleurs totalement dépourvu de connecteurs et de ponctuation (et notamment d'un point d'interrogation final), une majuscule à l'initiale :

- (30) J'ai mixer une grippe avec une bronchite et une crise d'asthme vous voyez ca souvent vous

La segmentation n'est pas non plus totalement absente dans l'exemple suivant, un connecteur (*et*) ouvrant la troisième et dernière phrase « syntaxique » :

- (31) trop heureuse pr miss Guyane elle est magnifique et elle a niqué en tte beauté languedoc la tchoin

En outre, il convient sans doute de relativiser l'absence fréquente de ponctuation finale, qui pourrait être rapportée à un « effet de genre », comme dans l'exemple suivant où aucune (autre) borne interphrastique ne manque :

- (32) Je suis toujours en avance sur mon temps. Par exemple, j'ai pas d'enfant et je regrette déjà d'en avoir fait. ça evite les conneries parfois

Une réflexion sur ce que recouvre, ou non, la sous-segmentation dans les tweets paraît tout particulièrement à mener à la lumière d'exemples comme celui qui suit, sans ponctuation et sans connecteur(s), néanmoins borné à gauche et à droite par respectivement trois ligateurs énonciatifs et un ponctuant (de même qu'à la lumière d'exemples comportant des émoticônes et/ou des émojis ; voir *infra*, partie 5) :

- (33) @X beh oui beh c pas lui qui va choisir pour moi hein

Par ailleurs, à l'inverse, certains tweets peuvent être considérés comme sur-segmentés. L'exemple suivant comporte

ce qui est parfois appelé un ajout après le point, ou encore un « complément différé », si l'on analyse *les arts martiaux* comme un complément d'objet direct « après coup » de l'élément central verbal *ont* (notons cependant que ce phénomène se rencontre dans les écrits littéraires) :

(34) @X Ils ont un coté décadent qu'il aime pas trop
mais également une morale et des us et coutumes
remarquables. Et les arts martiaux !

5. Les émoticônes et les émojis : des procédés de segmentation spécifiques dans les tweets (et autres écrits liés aux technologies numériques) ?

5.1. Quelques observations

Dans les tweets, comme dans des écrits d'autres genres liés aux technologies numériques, tels les SMS ou les courriels, le scripteur a la possibilité de recourir à des émoticônes et à des émojis. Ces signes iconiques sont des pictogrammes (au sens de Vaillant (2013)), visant la plupart du temps à indiquer l'émotion ou l'attitude énonciative du locuteur⁵. Un pictogramme se définit selon trois critères. D'abord, c'est un signe dont la saisie est iconique. Ensuite, c'est un signe qui appartient à un système d'écriture : il est compositionnel, constitué d'une combinaison d'unités iconiques minimales (Halté 2019), et il peut être placé (auprès de signes similaires, ou appartenant à d'autres systèmes sémiotiques) sur la chaîne syntagmatique et sur l'axe paradigmatique. Enfin, il faut que les pictogrammes, dans un système donné, aient la même taille, et soient dotés des mêmes couleurs, des mêmes formes, etc.

Nous appellerons « émoticône » tout pictogramme, constitué de signes du code ASCII⁶, qui indique (au sens

⁵ La terminologie utilisée pour désigner ces pictogrammes n'est pas fixée, plusieurs termes étant souvent employés : émoticônes, émojis, smileys, binettes, etc. La terminologie que nous proposons ici relève de choix méthodologiques organisant notre recherche, mais n'a pas de valeur dogmatique. Par ailleurs, nous choisissons de franciser le mot japonais *emoji* en « émoji », que nous accorderons au masculin.

⁶ *American Standardized Code for Information Interchange*, le tout premier système d'encodage de signes pour un usage informatique : lettres, chiffres, signes de ponctuation, etc.

sémantico-pragmatique du terme, celui de la *deixis*) l'émotion du locuteur ou, plus généralement, son attitude énonciative. On peut lire les pictogrammes de cette sorte en penchant la tête vers la gauche (émoticônes « occidentales » : :-) :-(:-P pour, respectivement, un sourire, une mimique triste et un tirage de langue) ou de face (émoticônes « orientales » : ^ _ O_o pour, respectivement, un sourire et une mimique de surprise)⁷.

Les cinq séquences d'émoticônes les plus fréquentes de notre corpus sont :

:) 0.009 par tweet
 ^^ 0.006 par tweet
 ;) 0.005 par tweet
 :(0.004 par tweet
 xD 0.004 par tweet

Nous appellerons « émoji » tout pictogramme dessiné accompagnant du texte dans les communications numériques, interagissant avec lui ou le remplaçant. Les émojis peuvent indiquer une émotion du locuteur, représenter un objet, une partie du corps ou un geste : 🍷 😊 😐 😓 😴 🐼...

Outre la différence formelle existant entre les deux signes, les émoticônes étant moins figuratives que les émojis, une autre différence est que ces derniers ne se réduisent pas à l'indication des émotions ou attitudes énonciatives du locuteur, contrairement aux émoticônes.

Les cinq séquences d'émojis les plus fréquentes sont :

😊 0.015 par tweet
 😊😊 0.005 par tweet
 😊 0.005 par tweet
 😊 0.004 par tweet
 😊😊😊 0.004 par tweet

⁷ Pour plus d'informations sur le sens des émoticônes et des émojis, nous renvoyons tout simplement à leurs pages *wikipedia* respectives. Nous n'avons pas la place, ici, d'aborder cette question.

Dans notre corpus, 11.42% des tweets se terminent par un emoji et 3.36% par une émoticône.

Les pictogrammes ont plusieurs fonctions dans les écrits numériques. Il est possible de les rapporter aux six fonctions discursives de Jakobson, comme le suggère par exemple Danesi (2013 : 100). Nous proposons de réduire ces fonctions à trois grandes catégories. Tout d'abord, les emojis, puisque certains d'entre eux servent simplement à représenter des objets, peuvent être utilisés pour remplacer un syntagme et ont alors des caractéristiques sémantiques et syntaxiques proches de celles d'un nom commun, d'un verbe, voire d'une proposition. Dans ce cas, ils ne servent pas à segmenter, puisqu'ils prennent tout simplement la place d'un élément syntaxique de la phrase, qu'il s'agisse d'un nom commun (ici pour le premier emoji) :

(35) Un grand 🙌 @X vous avez grave assuré toute vos chorégraphies vous êtes génial 👍. 🙌🙌🙌 pour votre super parcours #DALs ! 🍷🤔😊

Ou d'un verbe :

(36) (Okay je me lève malade, un mal de crâne pas possible, je bosse jusqu'à 18h (au lieu de 20h mdr) et j'organise une soirée chez moi. Je 😊).

Dans l'exemple (37), l'emoji se substitue même à une proposition :

(37) @X je t'aime bien mais 🙄

Nous laisserons ce cas de côté puisqu'il n'entre pas dans le cadre de la segmentation.

Une seconde fonction est « illustrative ». Dans ce cas, le pictogramme a aussi une fonction référentielle, mais il ne remplace pas un lexème. Il sert plutôt à illustrer, dans une relation de redondance sémantique (bien décrite par Klinkenberg (2009 : 26) concernant les rapports texte/image), une dénotation ou une connotation signifiée par la proposition :

(38) joue les pères Noël et vous accompagne jusqu'au réveillon ! 🎄

Troisième et dernière fonction : la modalisation (Halté 2018 : 59). Le pictogramme, dans ce cas, indique l'émotion ou l'attitude énonciative du locuteur. Il porte alors une modalité (au sens de Gosselin (2010)) servant à valider ou invalider une proposition. Les modalités signifiées par ces pictogrammes sont le plus souvent appréciatives et/ou épistémiques. C'est la fonction la plus utilisée : comme mentionné *supra*, les cinq émojis et émoticons les plus fréquents sont des icônes de mimiques faciales et indiquent donc une attitude ou une émotion du locuteur. Dans ce cas, le pictogramme employé vient modifier l'interprétation littérale d'une proposition, sur laquelle il fait porter l'attitude ou l'émotion du locuteur :

(39) **J'ai trop mangé chez ma pote, encore plus qu'hier 😊😊**...et elle m'a fait deux tupperware blindés pour chez moi (nous soulignons pour indiquer la portée)

Notons, enfin, que ces deux dernières fonctions peuvent occasionnellement s'hybrider. Ainsi, dans l'exemple suivant, le pictogramme représentant une tête d'extra-terrestre est à la fois illustratif (il est sémantiquement redondant avec le lexème *alien*) et modal (il indique l'émotion positive du locuteur, portant sur la proposition qui précède) :

(40) **Coupe vent Alien bien reçu 🛸** mais pq j'ai pas de stickers moi ? 😊 (nous soulignons)

Ces différentes fonctions conditionnent les lieux d'apparition des pictogrammes dans la chaîne syntaxique.

5.2. Où apparaissent les émoticons et les émojis dans la chaîne syntaxique ?

Danesi (2013 : 84) note que « [...] les émojis sont placés à des endroits qui sont habituellement occupés par des catégories de mots spécifiques, des marques de ponctuation, ou des particules discursives » (notre traduction). Danesi (2013 : 105) va même plus loin en rapprochant le fonctionnement des émojis « joyeux » de celui des virgules et des points : « La mimique joyeuse fonctionne habituellement comme une virgule ou un point dans les messages hybrides, ajoutant une valeur émotive aux pauses » (notre traduction).

Les mêmes remarques sont formulées, concernant les émoticônes, chez Dresner & Herring (2010). En effet, les émoticônes et les émojis, qu'ils soient employés comme illustrations ou comme modalisateurs, sont susceptibles d'apparaître aux mêmes endroits de la chaîne syntaxique que la ponctuation et les connecteurs. C'est notamment parce que leur fonction de modalisateur, alliée à leur dimension iconique, les rapproche de la ponctuation modale. Dürrenmatt (2015 : 23-24) évoque en effet, parmi les autres « opérations proposées au lecteur par la ponctuation », celle de la « modalisation », « la modalité [étant définie comme] l'expression de l'attitude du locuteur par rapport au contenu propositionnel de son énoncé ». Ainsi, à plusieurs égards, émoticônes et émojis peuvent être envisagés comme complémentaires, à l'écrit, des signes de ponctuation et des connecteurs, ce qui peut expliquer en partie leurs possibilités de positionnement dans la chaîne syntaxique.

Ils sont susceptibles (par ordre croissant de fréquence d'usage) d'être :

antéposés (3% des occurrences) :

(41) ❤️ Vacances à Bangkok à prix choc, pour une Saint Valentin loin d'être en toc !

en position interphrastique (apparition entre deux phrases « syntaxiques » au sein d'un même tour de parole, 19% des occurrences) :

(42) J'ai trop mangé chez ma pote, encore plus qu'hier
😬😬...et elle m'a fait deux tuperware blindés pour
chez moi

postposés (78% des occurrences) :

(43) Restons modestes, juste de gros connards 😬😬 !!!

Notons aussi que les émoticônes et les émojis apparaissant en tout début de tour de parole constituent souvent, lorsqu'ils sont modaux, une réaction à un contenu énoncé précédemment, et sont alors susceptibles de relever d'une forme de modalisation dialogique (Halté 2018 : 61).

De par ces caractéristiques positionnelles, ces pictogrammes « ont le plus souvent par ailleurs un rôle de démarcation des tours de parole mais peuvent intervenir à tout moment à la manière des points d'exclamation » (Dürrenmatt 2015 : 105).

Le rapprochement avec la ponctuation permet peut-être d'expliquer l'une des caractéristiques des pictogrammes étudiés ici : leur portée sémantique est quasi systématiquement dirigée vers le co-texte gauche (voir Halté (2017 : 9)), comme c'est le cas pour la ponctuation modale. Cette portée explique aussi la grande fréquence d'apparition des pictogrammes en toute fin des tours de parole. Ils peuvent donc marquer une segmentation en y ajoutant une valeur modale s'exerçant (le plus souvent) sur leur co-texte gauche, permettant ainsi de circonscrire un segment de la chaîne syntaxique et de faire porter sur lui une modalité particulière. Comme dans le cas de la ponctuation, il est possible pour les scripteurs d'agencer les pictogrammes en séquences plus ou moins répétitives, dont la fonction peut être d'intensifier la modalité indiquée ou de la complexifier. Ces séquences occupent les mêmes positions dans la chaîne syntaxique que les occurrences « simples ».

Il est fréquent que les émoticônes ou les émojis soient utilisés en remplacement pur et simple de la ponctuation, que ce soit au sein d'un tour de parole ou à sa fin, mais toujours de manière interphrastique :

(44)@X Je retire ce que j'ai dis alors ;) Je vous ai répondu sur Facebook !

Ils viennent alors souvent en place de signes de ponctuation (le locuteur, dans cet exemple, met une majuscule après l'émoticône comme s'il s'agissait d'un signe de ponctuation). Leur présence permet de relativiser une sous-segmentation souvent envisagée, traditionnellement, uniquement en fonction de l'emploi de signes de ponctuation et de connecteurs.

Par ailleurs, les pictogrammes, les connecteurs et la ponctuation, apparaissant nécessairement aux mêmes endroits de la chaîne syntaxique, peuvent aussi se combiner pour former des ensembles constituant des systèmes modaux plus ou moins complexes. Lorsqu'un pictogramme et un signe de ponctuation sont adjacents, tous les agencements sont possibles, que ce soit au milieu d'un tour de parole ou à sa fin. Ainsi, le pictogramme peut être placé avant le signe de ponctuation, ou après, et ce, qu'il s'agisse d'émoticônes (ici, avant le signe de ponctuation) :

(45)Pas le temps d'allumer les bougies que c'est déjà revenu :) .

Ou d'émojis (ici, après les signes de ponctuation) :

(46)Donc c'est ça la culture « geek» fin 2016??
😄😄😄😄 mais c'est désolant

Ou même de combinaisons entre signes de ponctuation, émoticônes et emojis :

(47)Mon allemand étant extrêmement limité tu me résumé cet article ^^ ?🤔

S'il est possible, notamment avec le point d'interrogation, que le pictogramme soit situé après le signe de ponctuation (comme dans l'exemple précédent), la combinaison

entre pictogrammes, ponctuation et connecteurs apparaît habituellement selon l'agencement suivant (nous soulignons) :

(48)@X je sais 😊, **mais** je vais essayer promis, juré, crach.. Non pas cracher !

Enfin, évoquons la proximité (illustrée et argumentée dans Halté (2018 : 133)) sémiotique, sémantique et syntaxique de ces pictogrammes avec les interjections, notamment avec les interjections acronymiques spécifiques des corpus numériques comme *lol* ou *mdr*. La segmentation à droite peut aussi être assurée par ces interjections, comme *ptdrrrrr* dans l'exemple suivant :

(49)@X Au four ????? azy jme casse ptdrrrrr

Là encore, les pictogrammes, la ponctuation, les connecteurs et les interjections peuvent se combiner dans des configurations déterminément agencées.

Conclusion

Nous espérons avoir précisé les procédés de segmentation à l'œuvre dans les tweets, écrits souvent spontanés et plus ou moins normés, et où se déploie donc une palette très riche d'outils : les signes de ponctuation de l'écrit, les connecteurs utilisés parfois comme dans les écrits monogérés et parfois comme dans les oraux polygérés, les ligateurs énonciatifs et les ponctuant de l'oral, les émoticônes, les émojis et les interjections des écrits numériques. Face à cette diversité, il pourrait être proposé la notion de segmenteur (corollaire de celle de segmentation), certains segmenteurs, tels les connecteurs, les ligateurs énonciatifs, voire les majuscules de début, étant orientés à droite et d'autres, tels les signes de ponctuation (notamment forte), les émoticônes, les émojis et les interjections, étant tendanciellement orientés à gauche.

Par la suite, dans le prolongement de ce travail sur la segmentation dans les tweets, il serait intéressant de se pencher plus particulièrement sur les connecteurs, ou encore sur les

combinaisons (contiguës) de segmenteurs. Une autre perspective pourrait aussi consister à dégager des profils de scripteurs, par exemple en établissant, avec un éclairage quantitatif, des corrélations entre l'utilisation de la segmentation et le niveau orthographique.

Références bibliographiques

- Abitbol J., Karsai M., Magué J.-P., Chevrot J.-P. & Fleury E. (2015). « Socioeconomic dependencies of linguistic patterns in Twitter: a multivariate analysis », *WWW '18 - World Wide Web Conference*, 1125-1134.
- Anis J. (1983). « Vilisibilité du texte poétique », *Langue française* 59 : 88-102.
- Béguelin M.-J. (dir.) (2000). *De la phrase aux énoncés : grammaire scolaire et descriptions linguistiques*. Bruxelles : De Boeck Duculot.
- Bessonnat D. (1991). « Enseigner... la ponctuation ?(!) », *Pratiques* 70 : 9-45.
- Bolander B. & Locher M. A. (2014). « Doing sociolinguistic research on computer-mediated data: a review of four methodological issues », *Discourse, Context and Media* 3(1) : 14-26.
- Bouchard R. (2001). « Alors, donc, mais... "particules énonciatives" et/ou "connecteurs" ? Quelques considérations sur leur emploi et leur acquisition », in G. Ledegen & N. Rossi-Gensane (éds) *Les grammaires du français et les « mots-outils »*. *Syntaxe & Sémantique* 3 : 63-73.
- Catach N. (1994). *La ponctuation*. Paris : Presses Universitaires de France.
- Catach N. (2003). *L'orthographe française. L'orthographe en leçons : un traité théorique et pratique*. Paris : Nathan (1^{ère} édition : 1995).
- Condamines A. (2006). « Avec et l'expression de la méronymie : l'importance du genre textuel », in G. Kleiber, C.

- Schnedecker & A. Theissen (éds) *La relation partie-tout*. Louvain-Paris : Peeters, 633-650.
- Danesi M. (2016). *The Semiotics of Emoji*. London-New York : Bloomsbury Academic.
- Dresner E. & Herring S. C. (2010). « Functions of the nonverbal in CMC: emoticons and illocutionary force », *Communication Theory* 20(3) : 249-268.
- Drillon J. (1991). *Traité de la ponctuation française*. Paris : Gallimard, coll. TEL.
- Dürrenmatt J. (2015). *La ponctuation en français*. Paris : Ophrys.
- Feuillard C. (1989). *La syntaxe fonctionnelle dans le cadre des théories linguistiques contemporaines*, Thèse d'État, Université Paris V.
- Gadet F. (2007). *La variation sociale en français*. Paris : Ophrys.
- Gadet F. (2007-2008). « L'oral et l'écrit dans les changements technologiques et idéologiques », in E. Galazzi & C. Molinari (éds) *Les français en émergence*. Berne : Peter Lang, 131-142.
- Gosselin L. (2010). *Les modalités en français. La validation des représentations*. Amsterdam-New York : Éditions Rodopi B. V.
- Halté P. (2017). « Positionnement syntaxique des interjections et des émoticônes : modalisation, portée, visée », *Cahiers de praxématique* 69. Consulté à l'adresse <https://journals.openedition.org/praxématique/4680>
- Halté P. (2018). *Les émoticônes et les interjections dans le tchat*. Limoges : Lambert Lucas.
- Halté P. (2019). « Iconicité et signification modale : l'émoticône, de l'icône du corps au geste énonciatif », *MEI* 47. Consulté à l'adresse <https://www.mei-info.com/en/revue/47/159/>
- Klinkenberg J.-M. (2009). « La relation texte-image. Essai de grammaire générale ». Consulté à l'adresse http://gemca.fltr.ucl.ac.be/docs/cahiers/20090128_Klinkenberg.pdf

- Koch P. & Oesterreicher W. (2001). « Langage parlé et langage écrit », in G. Holtus, M. Metzeltin & C. Schmitt (éds) *Lexikon der Romanistischen Linguistik*, vol I,2. Tübingen : Max Niemeyer Verlag, 584-627.
- Lodge A. (2011). « La question de la "langue commune" en français. Normes "sociales" vs normes "communautaires" », in S. Branca-Rosoff, J.-M. Fournier, Y. Grinshpun & A. Régent-Susini (éds) *Langue commune et changements de normes, Actes du Colloque International Langue commune et changements de normes*, Paris, février 2009. Paris : Champion, 77-92.
- Morel M.-A. & Danon-Boileau L. (1998). *Grammaire de l'intonation. L'exemple du français oral*. Paris : Ophrys.
- Paolacci V. & Rossi-Gensane N. (2012). « Quelles images de la phrase dans les écrits d'élèves de fin d'école primaire française ? Description linguistique et réponses didactiques aux difficultés des élèves », in F. Neveu, V. Muni Toke, P. Blumenthal, T. Klingler, P. Ligas, S. Prévost & S. Teston-Bonnard (éds) *Actes du 3^{ème} Congrès Mondial de Linguistique Française (CMLF)*, Lyon, juillet 2012, 341-359.
- Popin J. (1998). *La ponctuation*. Paris : Nathan.
- Roukhomovsky B. (2001). *Lire les formes brèves*. Paris : Nathan.
- Sébillot T. (1548). *Art poétique françoys*. Consulté à l'adresse <https://gallica.bnf.fr/ark:/12148/bpt6k50945f>
- Söll L. (1974). *Gesprochenes und geschriebenes Französisch*. Berlin : Schmidt.
- Trésor de la Langue Française informatisé* (TLFi), Centre National de Ressources Textuelles et Lexicales (CNRTL). Consulté à l'adresse <https://www.cnrtl.fr/definition/>
- Vaillant P. (2013). « Sémiologie des pictogrammes », *texto !* 18(4). Consulté à l'adresse <http://www.revue-texto.net/index.php?id=3336>

Annexe

Liste des émoticônes (298) :

[illegible]

Liste des émojis (90) :

😍	😭	👉	😘	😏	👍	💕	👏	😁	😞
😐	👋	😄	😱	🙌	😄	🙏	😓	💪	💕
😐	😏	😓	💕	😡	😄	👊	😓	😞	😭
😏	😭	😏	💜	💚	💕	✨	😓	💙	💋
🙏	😎	😏	😱	😏	😭	💕	😓	😄	😄
😏	😭	💕	😱	😏	😏	💛	😓	😭	👅
😏	😏	👹	😱	😏	😭	❤️	😏	🙏	🎉
🙏	🌸	✅	👑	😏	💔	👉	🐵	🌹	❤️
😏	😭	👍	👊	🔥	😏	😏	★	👉	😏